Downloaded from orbit.dtu.dk on: May 05, 2023

DTU Library

A systematic framework for enterprise-wide optimization: Synthesis and design of processing network under uncertainty

Quaglia, Alberto; Sarup, Bent; Sin, Gürkan; Gani, Rafiqul

Published in:
Computers & Chemical Engineering

Link to article, DOI:
10.1016/j.compchemeng.2013.03.018

Publication date:
2013

Link back to DTU Orbit

Citation (APA):
Quaglia, A., Sarup, B., Sin, G., & Gani, R. (2013). A systematic framework for enterprise-wide optimization: Synthesis and design of processing network under uncertainty. *Computers & Chemical Engineering*, *59*, 47-62. https://doi.org/10.1016/j.compchemeng.2013.03.018

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Computers and Chemical Engineering xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

A systematic framework for enterprise-wide optimization: Synthesis and design of processing networks under uncertainty

Alberto Quaglia [a], Bent Sarup [b], Gürkan Sin [a,*], Rafiqul Gani [a]

[a] CAPEC, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[b] Vegetable Oil Technology Business Unit, Alfa Laval Copenhagen A/S, 2860, Denmark

ARTICLE INFO

Article history:
Received 3 October 2012
Received in revised form 15 February 2013
Accepted 6 March 2013
Available online xxx

Keywords:
Enterprise-wide optimization
Integrated business and engineering
Decision-making under uncertainty
Mixed-integer non-linear program (MINLP)
Vegetable oil
Process synthesis
Product portfolio management

ABSTRACT

In this paper, a systematic framework for synthesis and design of processing networks under uncertainty is presented. Through the framework, an enterprise-wide optimization problem is formulated and solved under uncertain conditions, to identify the network (composed of raw materials, process technologies and product portfolio) which is feasible and have optimal performances over the entire uncertainty domain. Through the integration of different methods, tools, algorithms and databases, the framework guides the user in dealing with the mathematical complexity of the problems, allowing efficient formulation and solution of large and complex enterprise-wide optimization problem. Tools for the analysis of the uncertainty, of its consequences on the decision-making process and for the identification of strategies to mitigate its impact on network performances are integrated in the framework. A decomposition-based approach is employed to deal with the added complexity of the optimization under uncertainty. A network benchmarking problem is proposed as a benchmark for further development of methods, tools and solution approaches. To highlight the features of the framework, a large industrial case study dealing with soybean processing is formulated and solved.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The process industry sector is characterized by large capital investments, which are necessary for construction of production sites and facilities. The erection and commissioning of large production sites implies the use of massive amounts of economical, environmental and societal resources. The accuracy of the decision-making and of the design process is therefore of crucial importance, both for the enterprise which is committed to the investment, and for the human society in which the enterprise operates. Several tools have been developed and adopted in order to guide, support and facilitate the decision making process in capital investment projects, such as process management and project portfolio management (Project Management Institute, 2008).

Recent developments in Process Systems Engineering (PSE) have been focusing on formulating and solving processing network problems under the framework of enterprise-wide optimization (Grossmann, 2005). In this approach, the decision-making problem is cast in the form of superstructure optimization, which is formulated and solved as a Mixed Integer Linear or Mixed Integer Non Linear Programming (MIP or MINLP), making use of the integer and binary variables to represent discrete and binary choices.

The main strength of the enterprise-wide optimization approach is in its ability to provide comprehensive and transparent inputs to the decision makers, through a systematic and quantitative analysis. On the other hand, it poses several challenges, due to the size and complexity of the mathematical problem to formulate and solve, as well as to the amount of data which are required (Varma, Reklaitis, Blau, & Pekny, 2007). Often, the nature of the problem requires the formulation of large scale non-linear and non-convex problems (Karuppiah & Grossmann, 2006) whose solution to global optimality is still an open problem. Finally, the inclusion of data uncertainty in the decision-making problem causes a significant increase in problem size and complexity (Dua & Pistikopoulos, 1998; Karuppiah & Grossmann, 2008; Paules IV & Floudas, 1992; Sahinidis, 2004). Because of this complexity, formulation and solution of real industrial problems require considerable time and resources investment, as well as deep knowledge of optimization theory and algorithms.

For these reasons, we believe in the importance of developing of a systematic framework for enterprise-wide optimization particularly to motivate and facilitate its use in practice. The integration of state-of-art methods, tools and solution strategies, in a framework for enterprise-wide optimization has in fact the potential of increasing the productivity of the workflow needed to formulate and solve this class of problems; and thereby to enable the use of this powerful tool in industry and public sector, supporting transparent and efficient decision-making process.

* Corresponding author. Tel.: +45 45252806; fax: +45 45932906.
E-mail addresses: gsin@kt.dtu.dk, gurkansin@gmail.com (G. Sin).

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.03.018

Nomenclature

Indexes

i	component
k	process interval (origin)
kk	process interval (destination)
$react$	key reactant
rr	reaction
t	time (years)
s	Monte Carlo sample
f	first stage variable
s	second stage variable
det	solution of the deterministic problem

Parameters

MW_i	molecular weight
$P1_{i,kk}$	raw materials prices
$P2_{i,kk}$	utilities prices
$P3_{i,kk}$	products prices
$P4_i$	wastes disposal price
$SW_{i,kk}$	wastes fraction
$S_{k,kk}$	superstructure (binary)
$SP_{k,kk}$	superstructure of primary outlet (binary)
$\alpha^L_{j,kk}$	coefficient for capital cost estimation
$\beta^L_{j,kk}$	coefficient for capital cost estimation
$dist_{k,kk}$	transportation distance
$\alpha_{i,kk}$	fraction of utility mixed with process stream
$\gamma_{i,kk,rr}$	reaction stoichiometry
$SF_{i,kk}$	split factors
$\theta_{react,kk,rr}$	conversion of key reactant
$\mu_{i,kk}$	specific utility consumption
dr	discount rate
$F^{Max}_{i,kk}$	maximum throughput for interval kk
$Q^o_{j,kk}$	grid for piecewise linearization of throughput

Variables

$F_{i,k,kk}$	component *i* flow from process intervals *k* to process intervals *kk*
$ff_{i,kk}$	component flow after mixing
$R_{i,kk}$	utility flow
$F^M_{i,kk}$	component flow after mixing
F^{thr}_{kk}	throughput in interval *kk*
$F^{out1}_{i,kk}$	component flow leaving process intervals *kk* through primary outlet
$F^{out2}_{i,kk}$	component flow leaving process intervals *kk* through secondary outlet
$F^R_{i,kk}$	component flow after reaction
$Ctr_{k,kk}$	component flow after reaction
y_{kk}	selection of process intervals (binary)
$w_{j,kk}$	selection of an interval of the piecewise linearization (binary)
$Q_{j,kk}$	disaggregation variable for piecewise linearization of throughput

Operators

$E_\theta\ (f)$	expected value of function f over the domain of θ
P_s	probability of realization of event *s*

Abbreviations

CAPEX	capital investment
EBIT	earnings before interest and taxes
VSS	value of stochastic solution
UP	uncertainty price
EVPI	expected value of perfect information
UB	upper bound of the objective function
LB	lower bound of the objective function

In line with these considerations, in this manuscript we propose a systematic framework for synthesis and design of processing networks under uncertainty. The framework is based on the integrated business and engineering framework developed earlier (Quaglia, Sarup, Sin, & Gani, 2012a), which is extended to include decision-making under uncertainty.

The structure of the manuscript is as follows. In Section 1 the framework is described, by highlighting the mathematical formulation of the problem and the integration among the different methods and tools. In the Section 2 a Benchmark Network Problem (BNP) is proposed, and its formulation and solution according to the proposed framework is discussed. In the Section 3, the capability of the framework to deal with the size and complexity of an industrial problem is demonstrated, by formulating and solving a large scale case-study which is about synthesis of soybean processing network under uncertainty. Finally, conclusions and future works are presented in the last section.

2. The framework

In this section, the framework is presented in terms of its main components. A schematic representation of the framework is given in Fig. 1, where the integration of workflow, dataflow, solution methods and software tools are highlighted.

Step 1 (problem formulation) and 3 (deterministic problem) of the workflow correspond to our framework for synthesis and design of enterprise-wide processing network. A more extensive description of these steps can be found in (Quaglia et al., 2012a).

In the following section, each of the steps is described.

Step 1. Problem formulation

Problem definition

In this step the goal of the analysis is defined by stating the engineering, commercial and financial objectives of the project. On this basis, the objective function is selected and commercial and financial constraints such as maximum capital investment, success criteria among others are collected and systematized in terms of variable bounds.

Superstructure definition and data collection

All processing network alternatives including all possible raw materials, products and process technologies are generated and organized in a superstructure, constituted by a network of process intervals. A process interval is defined as a process section, which is able to perform a certain processing task. Commercial and engineering insights, as well as regulations are translated into logical constraints and variable bounds, to exclude infeasible networks from the search space. Relevant data for each of the process intervals are collected and organized in a predefined knowledge structure.

Model selection, development and validation

Models for each process interval contained in the superstructure are collected, or generated and validated if not available. Generic process interval models as described in earlier works (Quaglia et al., 2012a; Quaglia, Sarup, Sin, & Gani, 2012b) can be used here. An example of development of generic process interval model is reported in Appendix B. The superstructure, the interval

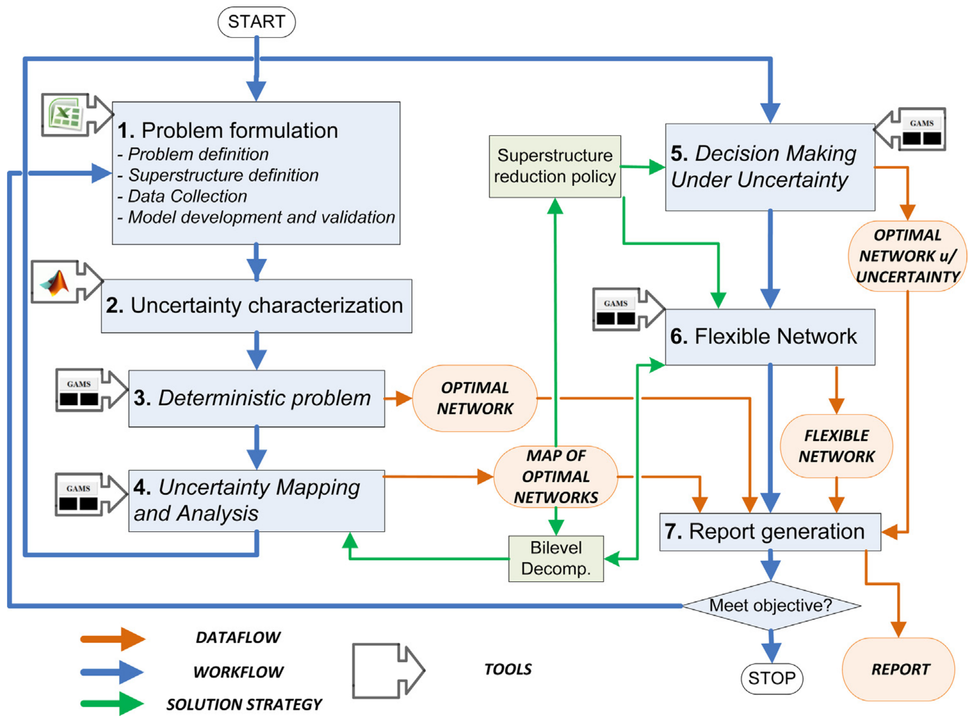

Fig. 1. The framework for decision-making under uncertainty, highlighted workflow, dataflow and tools.

models, the objective function, the logical constraints and the variable bounds defined in the previous steps constitute an MILP or MINLP model (Eqs. (1)–(6)).

$$\max f(x, y) \tag{1}$$

$$s.t. g(x, y) \geq 0 \tag{2}$$

$$h(x, y) = 0 \tag{3}$$

$$x \in X; \tag{4}$$

$$x^{LO} \leq x \leq x^{UP} \tag{5}$$

$$y \in \{0; 1\}^n \tag{6}$$

where f is the objective function, x represents the vector of continuous variables defined by their upper and lower bounds x^{UP} and x^{LO} in a continuous feasible region X, y is the vector of binary variables, g and h are the vectors of inequality and equality constraints respectively.

Step 2. Uncertainty characterization

In this step, the user is guided through the definition of the domain of uncertainty. Statistical analysis tool are integrated in this step in order to analyze the user-defined domain of uncertainty, and describe it in terms of Monte Carlo samples, which will be used for the mathematical formulation of the decision-making problem. Each of the uncertain data is described in terms of probability distribution (e.g. normal (Gaussian), uniform, Poisson, etc.), and its mean, maximum and minimum values are recorded in the input database. Then, the correlation between the uncertain data is analyzed, and characterized in terms of covariance between the probability distributions of each couple of data.

When available, experimental observations are used as input to the statistical analysis to characterize the uncertainties in data, e.g. mean and variance of an uncertain input data can be estimated from a sample of its observations if available. Unfortunately, for some of the data commonly used in enterprise-wide optimization problems (e.g. prices and market sizes of new products, performances of new processes, etc.), such observations are often not available. In these cases, an expert review approach is usually followed (Helton & Davis, 2003; Sin, Lantz, & Gernaey, 2009), in which the uncertainties in data are classified into several classes of uncertainty (e.g. low, medium and high) where each class of uncertainty has a predefined distribution with its variance and covariance.

Finally, the uncertain domain is sampled to generate a list of possible scenario with equal probability of realization, which will be used to approximate the solution of the probability integral. Uniform coverage of the uncertain space is obtained using Latin Hypercube Sampling; rank correlation method is used in order to reflect the correlation between the uncertain parameters in the generated future scenarios (Iman & Conover, 1982).

Step 3. Deterministic formulation and solution

In this step, the uncertain data are fixed at their mean value, and the enterprise-wide optimization problem defined by Eqs. (1)–(6) is solved. The results indicate the optimal processing network in deterministic conditions, by selecting the raw materials, the product portfolio and the processing path, as well as the value of all variables free for optimization. The value of the objective function and other performance indicators at optimality are reported.

Step 4. Uncertainty mapping and analysis

For each of the scenarios generated by sampling from uncertainty domain in step 4, a deterministic enterprise-wide optimization problem is solved. The result is a distribution of optimal processing networks and of objective function values obtained for different realization of uncertain data.

These distributions are analyzed to infer about the consequences of the uncertainty on the decision-making process. For example, the number of optimal networks which are identified for different values of the uncertain parameters and the variability of the objective function gives an indication of the magnitude of the impact of data uncertainty on the decision-making process.

Step 5. Decision making under uncertainty

In this step the problem of synthesis of processing network under uncertainty is formulated as a stochastic programming problem (Eqs. (7)–(13)), where the expected value of the objective function over the uncertain domain is maximized.

$$\max E_\theta[f(x,y,\theta)] \tag{7}$$

$$s.t. g(x,y,\theta)\geq 0 \tag{8}$$

$$h(x,y,\theta)=0 \tag{9}$$

$$x\in X \tag{10}$$

$$x^{LO}\leq x\leq x^{UP} \tag{11}$$

$$y\in\{0;1\}^n \tag{12}$$

$$\theta\in\{\theta^{LO};\theta^{UP}\}^m \tag{13}$$

where θ is the vector of uncertain data and $E_\theta(f)$ is the expected value of the objective function f over the θ space.

The calculation of the expected value of the objective function requires the evaluation of a multidimensional integral. A common approach in stochastic programming consists of using a Monte Carlo sampling based method for the approximation of the expected value of the objective function, generally known as Sample Average Approximation (SAA) (Eq. (14)).

$$E_\theta[f(x,y,\theta)]\sim\sum_{s=1}^{NS}P_s\cdot f(x,y,s) \tag{14}$$

$$s.t. g_s(x,y,s)\geq 0,\quad \forall s\in S \tag{15}$$

$$h_s(x,y,s)=0,\quad \forall s\in S \tag{16}$$

where NS is the number of Monte Carlo samples in the sample matrix S, s is one Monte Carlo sample and P_s the probability of realization of sample s.

It is important to note that, in this formulation the number of equations is increased, because each constraint containing an uncertain data has to be rewritten for each sample s (Eqs. (15)–(16)).

In the framework, the solutions of previous steps are used for variable initialization and bounding, in order to manage the increased numerical size and complexity of the problem under uncertainty.

The results identify the network which is feasible over the entire uncertain space and whose expected value of objective function is optimal. This solution is called optimal network under uncertainty.

It is important to remember that the maximization of the expected value corresponds to risk-neutral decision-making (Shapiro, 2012). Other conditional values (such as, for example, value-at-risk) can be used instead of the expected value, to reflect different attitude with respect to risk.

Step 6. Flexible network

In many cases, the exact realization of the uncertain data is unknown in the design phase, but becomes known in the operational phase. As an example, let us consider the price of a certain product: while designing a processing network, this data is an estimation of the average over the entire investment horizon (usually between 10 and 25 years). As a consequence, a high degree of uncertainty is associated with this data (i.e. the standard deviation of its probability distribution function has a large value). While operating the network however, this price becomes an observable data hence becomes known implying reduced uncertainty.

Because of the additional knowledge available at operational stage (when some of the uncertain data become observable) the operational policy defined in the design stage may become suboptimal, and a new optimal operational policy exists. The problem of identifying the optimal operational policy with respect of production volumes, patterns, etc. for an existing facility or network (i.e. within the flexibility allowed by a given design) is defined as planning problem (Erdirik-Dogan & Grossmann, 2008; Ierapetritou, Pistikopoulos, & Floudas, 1996; Terrazas-Moreno & Grossmann, 2011).

It is evident that, the higher the flexibility allowed by the design of a network, the higher is the possibility to modify the operational policy and to adapt to a different realization of the uncertain data, once these become observable. At the same time, higher design flexibility is often obtained at the expense of a larger capital investment (for example through redundant design or overdesign). As a consequence, a trade-off exists between the capital investment and the ability of mitigating negative consequences of the uncertainty at planning and operational stage.

In the industrial design practice, the desired level of flexibility of the design is determined on the basis of previous experience, heuristic or safety factors. In this work, we attempt to identify the optimal trade-off between flexibility and capital investment (here defined as the *optimal flexible network*) through explicit consideration of the uncertainty while solving the synthesis and design problem.

To identify the optimal flexible network, design and optimization variables are partitioned into first and second stage variables, and the problem of decision-making under uncertainty is reformulated as a two stage stochastic programming with recourse. In Eqs. (17)–(24) the deterministic equivalent formulation of the problem is presented (Birge & Louveaux, 1999).

$$\max f_f(x_f,y_f)+E_\theta[f_S(x_s,y_s,\theta)] \tag{17}$$

$$s.t. g(x_s,y_s,\theta)\geq 0 \tag{18}$$

$$h(x_s,y_s,\theta)=0 \tag{19}$$

$$p(x_f,y_f,x_s,y_s,\theta)\geq 0 \tag{20}$$

$$q(x_f,y_f,x_s,y_s,\theta)=0 \tag{21}$$

$$x^{LO}\leq x\leq x^{UP} \tag{22}$$

$$y\in\{0;1\}^n \tag{23}$$

$$\theta\in\{\theta^{LO};\theta^{UP}\}^m \tag{24}$$

where the subscript f indicates first stage variables and s second stage variables.

First stage variables corresponds to design decisions, which have to be taken before having full knowledge of the realization of the uncertain data; these include the binary variables indicating the selection of process intervals and the maximum design flows through each interval. The second stage variables are operational decisions, which can be fixed after the realization of the uncertainty, and include material flows and operating conditions. In the adopted formulation of the stochastic decision problem, the second

Table 1
Decision-making under uncertainty indicators.

Name	Definition	Interpretation
EPVI	$E(\max f(x,y,\theta)) - \max E(f(x,y,\theta))$	Maximum price which is fair to pay to have perfect information on the uncertain data
VSS	$\max E(f(x,y,\theta)) - E(f(x^*_{det}, y^*_{det}, \theta))$	Value of performing the stochastic analysis
UP	$\max(f(x,y)) - \max(E(f(x,y,\theta)))$	Price associated to uncertainty

stage decision variables are explicitly described for all scenarios (Birge & Louveaux, 1999).

In order to link design (or first stage) and operational (or second stage) variables, constraints p and q (Eqs. (20)–(21)) are added. These additional constraints define boundaries for the value of the second stage decisions as a function of the value of the first stage variables. For example, the flow through a process or equipment (second stage variable) has to be less or equal than the maximum design flow for that process or equipment (first stage variable).

Step 7. Report generation

In the last step a report containing all the relevant information generated by the analysis is prepared. This includes the optimal network structures obtained at different steps, and the performance metrics calculated at optimality. Moreover, indicators relevant to the decision-making are calculated from the above mentioned results. These indicators include expected value of perfect information (EVPI), value of stochastic solution (VSS) and uncertainty penalty (UP). A definition of these indicators, which are calculated both for the optimal network under uncertainty and for the optimal flexible network, together with their interpretation is reported in Table 1 (Birge & Louveaux, 1999).

The EVPI calculates the expected value of objective function loss which is caused by not knowing the exact value of the uncertain data. Alongside with its importance for the design process, this indicator gives valuable insights also from a process management point of view, by providing a measure of the maturity of the project. For example, a large value of EVPI suggests keeping the project in the exploratory phase (also called Front End Loading in process management terminology) and trying to reduce data uncertainty, since it promises to largely benefit the performances of the network. Moreover, the EVPI gives an indication of the upper bound of the cost which such an exploratory phase should have.

The VSS is the difference in performances between the implementation of the stochastic and the deterministic solutions, both evaluated in uncertain conditions. It quantifies the value of obtaining the solution under uncertainty.

The UP quantifies the reduction in performances which is associated to the necessity of considering uncertainty in the data, and therefore provides an estimation of the cost of the uncertainty.

These indicators, presented as a package together with the solution of the previous steps, constitute a comprehensive, transparent and quantitative input to decision-makers, enabling them to take good decisions through a rational decision-making process.

2.1. Managing the complexity through the framework – solution strategy

In the previous sections, the workflow for the synthesis of processing networks has been presented as a succession of steps, in which different optimization problems are formulated and solved, and the result analyzed and processed to provide insights for the decision-making process.

When dealing with large industrial problems, the solution of these optimization problems may require large computational time and resources, especially when a large number of uncertain data is considered (Sahinidis, 2004).

In order to allow the solution of large scale industrial problems, our framework integrates a decomposition-based solution approach. In the next sections, the algorithms and solution strategy employed are described.

2.1.1. Superstructure reduction policy

As previously explained, problems of decision-making under uncertainty are solved through sampling-based methods. This method results in the increase of problem size, which may cause the problem to become difficult to solve.

Because of the incremental structure of the workflow integrated in our framework, a large amount of information and results are generated prior to the formulation of the stochastic problem. This information and intermediate results are used as a rationale for the simplification of large complex problems to a solvable form.

An example of a simplification strategy based on intermediate results is the *superstructure reduction policy*, which is integrated in our framework. The superstructure reduction policy consists of elimination of the process intervals which have not been selected in any of the uncertainty mapping solutions. As a result, the problem under uncertainty is formulated and solved for a smaller superstructure, which implies a reduction in terms of problem size (number of constraints and binary variables) and of search space, and consequently of computational resources necessary for the solution of the problem.

It should be underlined that the superstructure reduction policy is a simplification scheme obtained under a user-defined domain of uncertainty, and therefore the user must be aware of its effect on the quality of the obtained solution. For a general optimization problem, the superstructure reduction policy is not an exact method, and the solution of the simplified problem may not be the global optimum of the original problem. As a consequence, the superstructure reduction policy should be applied only for problems which are otherwise not tractable, and for which an increased gap for global optimality is acceptable.

Observation 1. Under the condition of (i) problem linearity, (ii) uniform probability distribution of the uncertain data, (iii) feasibility and optimality of the same base for every network over the uncertain domain; the proposed superstructure reduction policy is an exact method, and the solution of the simplified problem is the global optimum solution of the original stochastic problem.

A proof of observation 1 is given in Appendix A.

2.1.2. Bi-level decomposition

Bi-level decomposition is a commonly used solution scheme for optimization problems. (Erdirik-Dogan, Grossmann, & Wassick, 2007; Kopanos, Puigjaner, & Georgiadis, 2009; Terrazas-Moreno & Grossmann, 2011).

This solution strategy is based on the decomposition of the problem into upper and lower level problems. The former is a relaxed version of the original problem, which is solved to determine an upper bound (in case of maximization problems) of the objective function. The latter is obtained by fixing the value of some decision variables to the solution of the upper level, to obtain a lower bound of the objective function. The 2 problems are solved iteratively (cuts are added at each iteration to exclude previous solutions from the search space) until a stopping criterion is met. The effectiveness of

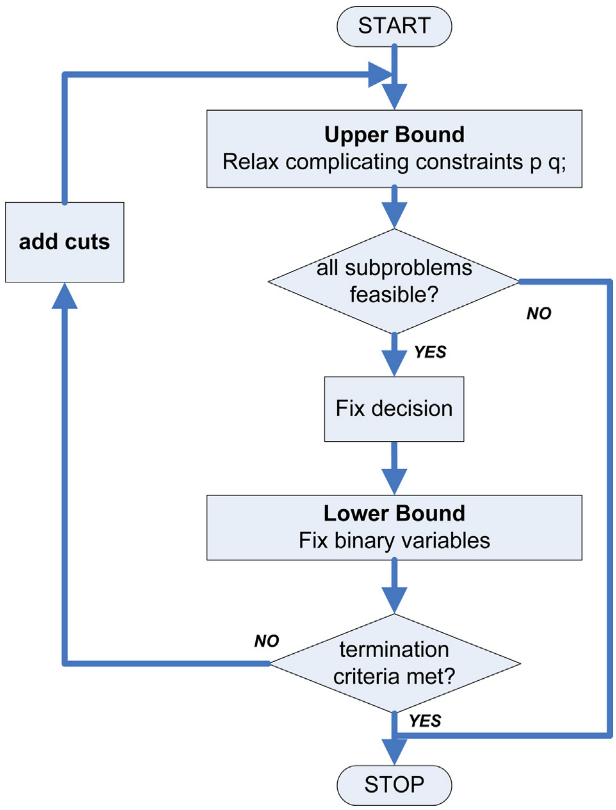

Fig. 2. Bi-level decomposition procedure adapted after Erdirik-Dogan et al. (2007).

this solution approach is highly dependent on the tightness of the relaxation used to define the upper level problem, and in the worst case it may require the complete exploration of the solution space (Erdirik-Dogan et al., 2007).

A bi-level decomposition strategy is developed and integrated in our framework for the solution of the flexible network problem. The solution algorithm is reported in Fig. 2.

The upper bound (for maximization) is obtained by relaxing the complicating constraints p and q (Eqs. (20)–(21)) and solved. As highlighted in the incidence matrix in Fig. 3, because of its structure the relaxed problem can be further decomposed into a series of smaller sub-problems, which can be easily solved in a specific sequence (Ierapetritou, Acevedo, & Pistikopoulos, 1996). It should be noted that, if this second decomposition is performed, the resulting sub-problems are equivalent to the one solved in the uncertainty mapping step (step 4).

The binary variables (representing network topology) are then fixed at the solution of the upper bound problem, and the lower bound problem is solved. If any of the termination criteria (gap, number of iterations, trend of the lower bound solution, etc.) is met, the procedure is stopped; otherwise cuts are added to exclude the previous solution and the procedure is repeated. The solution is reported together with the termination criterion which has been met.

3. Numerical example: network benchmark problem

A network benchmark problem (NBP) is proposed to test and demonstrate the features of the framework. NBP data and models are reported in Appendix C, to allow the interested reader to completely reproduce the problem and its solution, as well as to use it for benchmarking purposes. In the following sections, the application of the framework to the solution of the NBP is described.

Step 1. Problem definition

The objective of NBP is the synthesis of the processing network which maximizes the Earnings Before Interest and Tax (EBIT). The superstructure is composed of 2 possible raw materials, 6 process intervals organized in 3 processing steps, and 4 potential products.

The process interval models are based on the generic process interval model described in (Quaglia et al., 2012a). Model equations are reported in Appendix C.

Step 2. Uncertainty characterization

Five data are considered uncertain in the NBP: 2 raw material prices, 1 product price and 2 raw material compositions. Since NBP is a conceptual problem, the probability distribution and correlation information are assumed to be known a priori from observations. In Table 2, the probability distribution for each uncertain data is described by defining the probability distribution function, as well as maximum and minimum value. In Table 3, the correlation between the uncertain data is described in terms of covariance between each data pair. As an example, the covariance between the two raw material price data ($P1_{I-1}$ and $P1_{I-2}$) is defined as 0.8, indicating the existence of a strong positive correlation between these data.

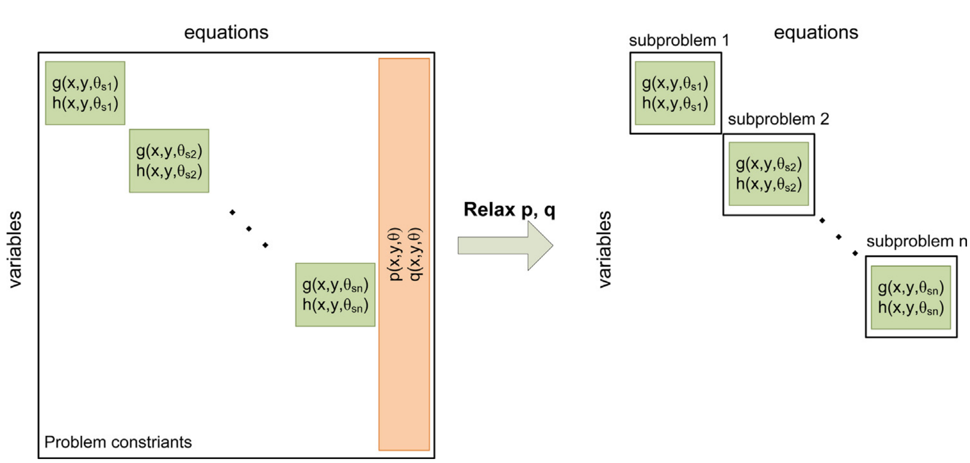

Fig. 3. Incidence matrix for the optimization problem. On the left the structure of the original problem under uncertainty, on the right the upper bound problem (for maximization).

Table 2
Probability distribution of the uncertain data.

Data	Mean	Probability distribution	Min	Max	Description
P1(I-1)	16	Uniform	13.92	18.08	Raw material I-1 price
P1(I-2)	18	Uniform	16.92	19.08	Raw material I-2 price
P3(V-3)	70	Uniform	52.50	87.50	Product V-3 price
φ(C-2,I-1)[a]	15	Uniform	10.50	19.50	Raw material I-1 component C-2 flow
φ(C-2,I-2)[a]	20	Uniform	8.00	32.00	Raw material I-2 component C-2 flow

[a]Variation of C-2 flow are compensated by opposite variation of C-3 flow.

Table 3
Covariance matrix.

	P1(I-1)	P1(I-2)	P3(V-3)	φ(C-2,I-1)	φ(C-2,I-2)
P1(I-1)	**1**	0.8	0.4	−0.4	−0.4
P1(I-2)	0.8	**1**	0.4	−0.4	−0.4
P3(V-3)	0.4	0.4	**1**	0	0
φ(C-2,I-1)	−0.2	0	0	**1**	0.4
φ(C-2,I-2)	0	0	0	0.4	**1**

The uncertain parameter space is sampled to define 200 future scenarios, reflecting the defined characterization of the uncertainty domain. Fig. 4 shows how the positive correlation between the raw material price data is maintained in the generated samples.

Step 3. Deterministic formulation and solution

The uncertain data are fixed at their expected values and a deterministic problem with 2261 constraints and 120 binary variables is formulated and solved in GAMS, using CPLEX solver (IBM Corp., 2009). The optimal network is identified in the results, and an EBIT value of 4038.826 k$/year is calculated.

Step 4. Uncertainty mapping

For each of the 200 future scenarios generated in step 2, a separate optimization problem is formulated and solved. The results identify 6 different networks as a function of the uncertain data realization (see Table 4 and Fig. 5). A large variability of objective function value is observed (from 2399.5 to 5846.9 k$/year). Two process intervals (II-3 and IV-1) are never part of an optimal network.

All these factors indicate that the uncertainty in the data has a large impact on the decision making and on the performances of the network, and therefore needs to be carefully considered in the decision-making process.

Step 5. Decision making under uncertainty

The problem under uncertainty is formulated and solved using sample average approximation (SAA). The resulting MILP problem is constituted by 398,669 constraints and 120 binary variables. The optimal network under uncertainty (see Fig. 6) is different from the optimal solution obtained under deterministic conditions. The calculated EBIT is now 2778.089 k$/year.

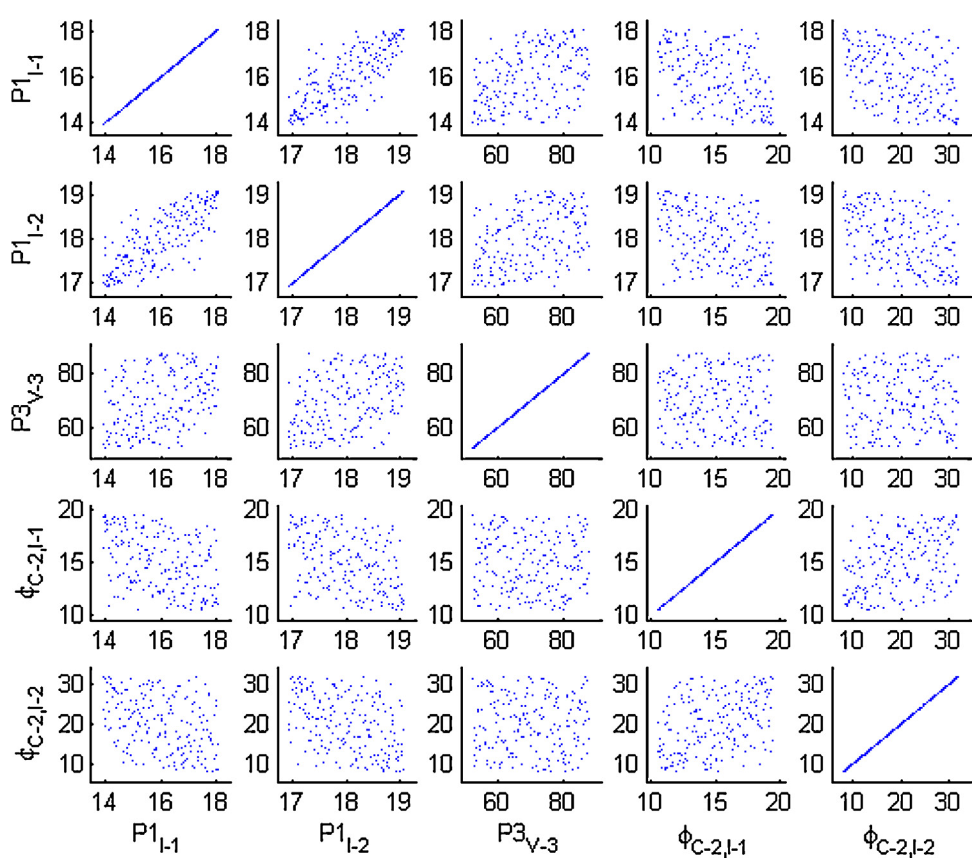

Fig. 4. Latin hypercube samples with correlation control from uncertainty domain.

Table 4
Uncertainty mapping results – selected networks.

Network	Probability	Selected intervals											
		I-1	I-2	II-1	II-2	II-3	III-1	III-22	IV-1	V-1	V-2	V-3	V-4
1	53.5%	1	0	0	1	0	0	1	0	0	0	1	0
2	27.0%	1	0	1	0	0	1	0	0	1	1	0	0
3	8.5%	1	0	1	0	0	0	1	0	0	0	1	0
4	7.0%	0	1	0	1	0	0	1	0	0	0	1	0
5	3.0%	0	1	1	0	0	1	0	0	1	1	0	0
6	1.0%	0	1	0	1	0	1	0	0	1	1	0	0

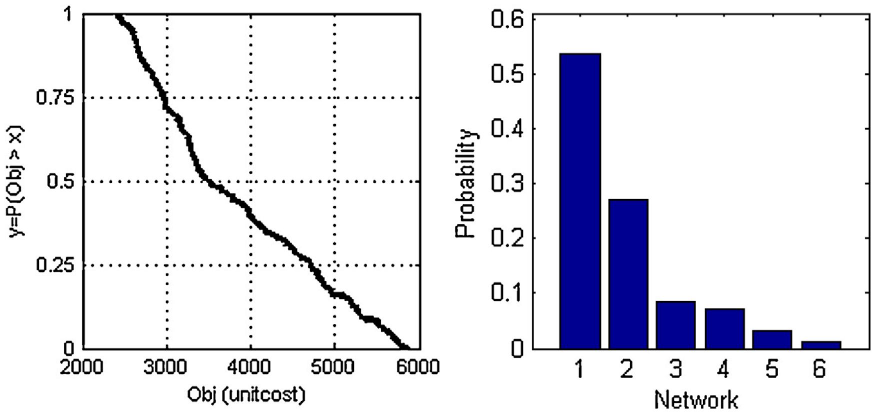

Fig. 5. Uncertainty mapping results. On the left the cumulative distribution of objective function values, on the right the distribution of optimal networks.

Step 6. Flexible solution

The 2 stage stochastic programming with recourse is formulated and solved using the proposed bi-level decomposition approach. The solution is obtained in 3 iterations and the solver stops because of deterioration of the upper bound (Table 5).

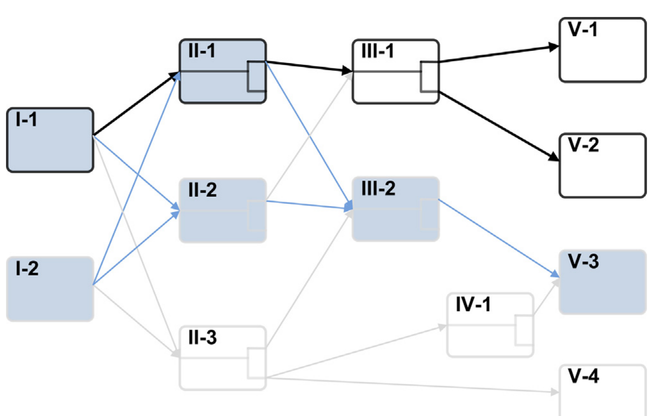

Fig. 6. NBP solution. In bold, the optimal network under uncertainty (step 5). In blue, the flexible network (step 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Solution of the flexible network problem through bi-level decomposition.

Iteration	Bounds	Topology
Iteration 1	$UB_1 = 3736.408$ $LB_1 = 2288.271$	I-1; I-2; II-1; II-2; III-1; III-2; V-1; V-2; V-3
Iteration 2	$UB_2 = 3508.548$ $LB_2 = 3099.514$	I-1; I-2; II-1; II-2; III-2; V-3
Iteration 3	$UB_3 = 2891.587$ $LB_3 = 2068.667$	I-1; I-2; II-1; II-2; III-1; III-2; V-1; V-2

The resulting optimal flexible network is reported in Fig. 6. The optimization selects a redundant structure for the first two steps, for which 2 intervals per step are selected. Such a redundancy allows a greater operational flexibility, which can be exploited at a planning stage once additional knowledge on the uncertain data becomes available, at the expense of an increase of capital cost. The expected value of the EBIT is calculated as 3099.514 k$/year, which corresponds to an increase of more than 11% with respect to the optimal network under uncertainty. Such an improvement is obtained thanks to the integration of operational flexibility considerations in the design decision-making phase.

Step 7. Report generation

As a final step, a report containing the above described results and the indicators is generated, providing an overview of all the obtained information (Table 6).

Three different networks are obtained as deterministic, under uncertainty and flexible solution.

The EVPI estimates in 958.319 k$/year (24% of the EBIT) the performance gain which could be obtained by knowing the exact value of the uncertain data. As previously discussed, this indicator provides also a measurement of project maturity, and can be used for process management decision (along with other consideration, such as time-to-market, company strategy etc.).

Table 6
Results report for the benchmark example problem.

Solution	Network	EBIT (k$/year)
Optimal network	I-1; II-2; III-2; V-3	4038.826
Network u/uncertainty	I-1; II-1; III-1; V-1; V-2	2778.089
Flexible network	I-1; I-2; II-1; II-2; III-2; V-3	3099.514

Indicator	VSS (k$/year)	UP (k$/year)	EVPI (k$/year)
Network u/uncertainty	633.529	1260.737	958.319
Flexible network	954.955	939.312	

Table 7
Probability distribution of the uncertain data.

Data	Mean	Probability distribution	Min	Max	Description
P1(1)	350.00	Uniform	245.00	455.00	Soybean low quality price
P1(2)	355.00	Uniform	248.50	461.50	Soybean high quality price
P2(Steam-10b)	25.00	Uniform	18.75	31.25	Steam-10b price
P3(52)	900.00	Uniform	630.00	1170.00	Refined oil price
P3(60)	310.00	Uniform	155.00	465.00	Low pro meal price
P3(62)	320.00	Uniform	272.00	368.00	High pro meal price
P3(63)	900.00	Uniform	765.00	1035.00	Protein concentrate price
φ(1,protein)	38.00	Uniform	34.20	41.80	Protein content low quality soy
φ(2,protein)	38.00	Uniform	36.10	39.90	Protein content high quality soy
μ(steam-10 bar,36)	0.33	Uniform	0.27	0.40	Steam consumption protein separation
Split(11,NHP)	0.50	Uniform	0.43	0.58	Separation NHP water degumming

The impact of data uncertainty on the investment is indicated by the value of the stochastic solution (VSS) and the uncertainty price (UP). The large values of the former indicator show how the performances of the deterministic solution decline when data uncertainty is considered, and therefore underline the importance of considering data uncertainty in the decision-making process.

The UP shows the decay in financial performances of the investment, due to the need of considering uncertainty in the decision making. The effectiveness of the flexible network in making use of network flexibility to mitigate the consequences of the uncertainty is shown by the reduction in UP (−25%) with respect to the optimal network under uncertainty. On the base of these results, the analysis suggests the flexible network as the best decision.

4. Case study – soybean oil processing

In order to demonstrate the ability of the proposed framework to manage problems of industrial complexity, an industrial case study dealing with the synthesis and design of a soybean processing network is formulated and solved. This case study (developed in collaboration with Alfa Laval) has been discussed in our earlier works (Quaglia et al., 2012b, 2012a), where problem formulation, deterministic solution, uncertainty mapping and solution under uncertainty (corresponding to steps 1–5 of the framework) have been presented; for this reason, the next section will focus mostly on step 6 and 7. For more detailed information on the other steps, the reader is invited to consult the two references given above. To protect the confidentiality of the data, all the solutions are here reported in scaled units.

Steps 1–5

The problem deals with the synthesis of a soybean processing network. A superstructure composed by 2 raw materials of different quality, 42 process technologies, 20 potential products and 1 special waste is generated. The maximization of the EBIT over a 25 years period is selected as objective function and the generic process interval model is adopted. The resulting discrete non-linear problem is simplified to a linear optimization one through piecewise linearization of the capital cost constraint (Floudas, 1995), resulting in the formulation of a MILP problem with 176,000 constraints and 485 binary variables.

Eleven data are identified as uncertain (7 of which are related to the market conditions, 2 to the composition of the raw materials and 2 to process performances). Data distributions and correlation coefficients have been obtained through statistical analysis of historical data, when available (for standard quality soybean price, oil price, low-protein meal price, high protein meal price, through the Chicago Board of Trade). Expert review has been used for the remaining data such as raw material composition and process performances.

The user-defined uncertainty domain is described in Tables 7 and 8, where the probability distributions of the uncertain data and their correlation are reported.

The deterministic problem is solved and the optimal network identified. Uncertainty mapping is performed and 6 optimal networks are identified for different future scenarios. A stochastic problem is formulated and solved to identify the optimal network under uncertainty (see Table 9).

Table 8
Covariance matrix.

	Soybean low quality price	Soybean high quality price	Steam-10b price	Refined oil price	Low pro meal price	High pro meal price	Protein concentrate price	Protein content low quality	Protein content high quality	Steam protein separation	Separation NHP water degumming
Soybean low quality price	**1.00**	0.58	0	0.56	0.30	0.38	0.53	0.38	0	0	0
Soybean high quality price	0.58	**1.00**	0	0.41	0.35	0.32	0.58	0	0.56	0	0
Steam-10b price	0	0	**1.00**	0	0	0	0	0	0	0	0
Refined oil price	0.56	0.41	0	**1.00**	0.78	0.66	0	0	0	0	0
Low pro meal price	0.30	0.35	0	0.78	**1.00**	0.74	0	0	0	0	0
High pro meal price	0.38	0.32	0	0.66	0.74	**1.00**	0	0	0	0	0
Protein concentrate price	0.53	0.58	0	0	0	0	**1.00**	0	0	0	0
Protein content low quality	0.38	0	0	0	0	0	0	**1.00**	0	0	0
Protein content high quality	0	0.56	0	0	0	0	0	0	**1.00**	0	0
Steam cons. protein separation	0	0	0	0	0	0	0	0	0	**1.00**	0
Separation NHP water degumming	0	0	0	0	0	0	0	0	0	0	**1.00**

Please cite this article in press as: Quaglia, A., et al. A systematic framework for enterprise-wide optimization: Synthesis and design of processing networks under uncertainty. *Computers and Chemical Engineering* (2013), http://dx.doi.org/10.1016/j.compchemeng.2013.03.018

Table 9
Results statistics.

Case study	NBP			Soybean
Problem	Deterministic	Under uncertainty	Flexible	Flexible
No. binary variables	120	120	2508	485
No. constraints	2260	398,669	609,910	12,296,000
Relative optimality tolerance	1E-6	1E-6	1E-6	8.3%
CPU time (s)	0.33	56.12	316.41	106,360
Solution algorithm	Direct	Direct	Bilevel decomp.	Bilevel decomp.
No. outer iterations	–	–	4	16
Average CPU time for 1 iteration (s)	–	–	79.10	4432

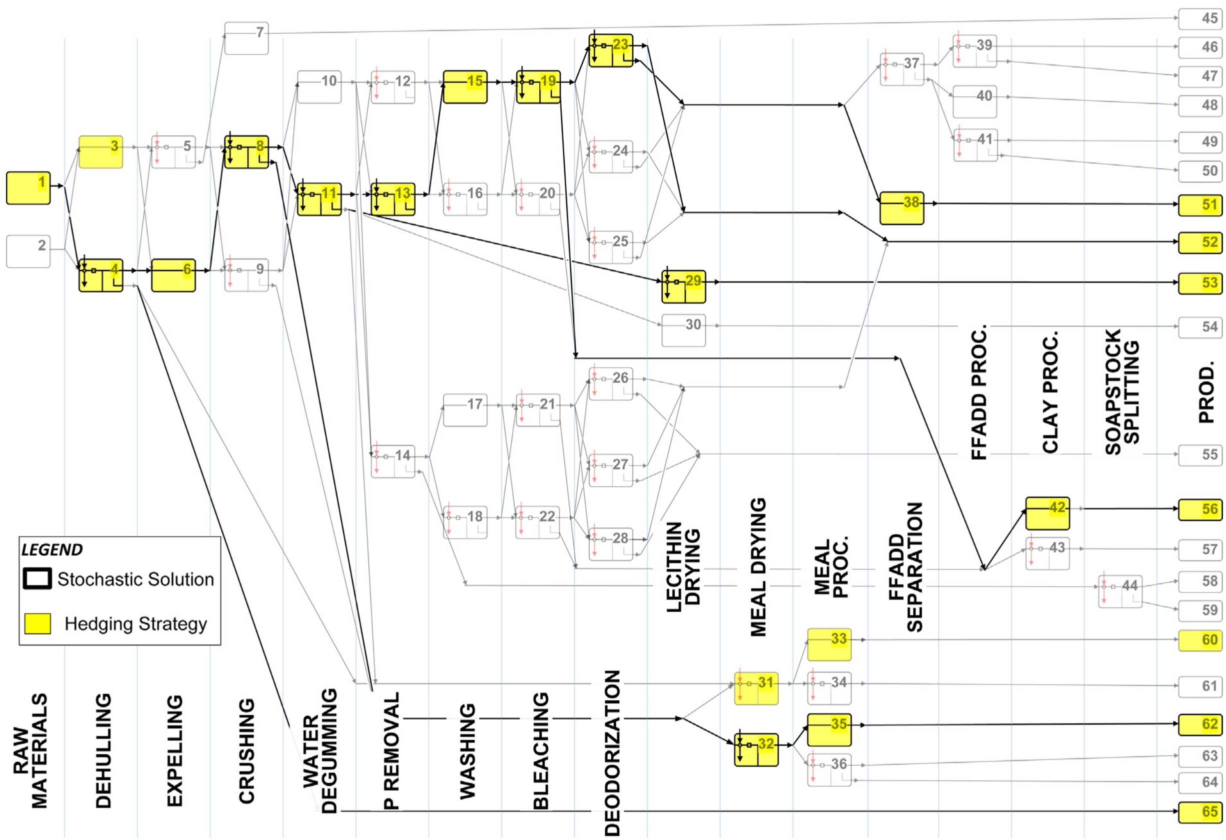

Fig. 7. Soybean oil processing network solution. In bold the optimal stochastic solution, in yellow the flexible network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Step 6. Flexible network solution

The flexible network problem is formulated as described in the previous section, resulting in the formulation of a MILP problem with more than 12 million constraints. Because of the size of the problem, its direct solution poses extreme computational challenges.

Therefore, the above described superstructure reduction policy is used to simplify the problem to a tractable form. Since the conditions for observation 1 are not satisfied, this simplification is not an exact method.

The simplified problem is solved using the bi-level decomposition strategy. Because of the size and complexity of the problem, the convergence criterion is set to accept a gap of 10%. Problem statistics are reported in Table 9. The solver stops because the convergence criterion is satisfied after 16 iterations and a total computational time of 29 h (on a standard computer with CPU Intel Core^{TM} i5 2.53 GHz). Although large, the computational time

Table 10
Soybean processing results report.

Solution	Network		EBIT (unit-cost/year)
Optimal network	1 4 6 8 11 13 15 20 23 29 32 35 38 42 51 52 53 56 62 65		108.5
Network u/uncertainty	1 4 6 8 11 13 15 19 23 29 32 35 38 42 51 52 53 56 62 65		106.8
Flexible network	1 3 4 6 8 11 13 15 19 23 29 31 32 33 35 38 42 51 52 53 56 60 62 65		115.3
Indicator	VSS (unitcost/year)	UP (unit-cost/year)	EVPI (k$/year)
Network u/uncertainty	1.8	0.3	19.3
Flexible network	8.8	−6.7	

required for the solution is compatible with the requirements for a design problem. Further reduction of the computational time can be obtained by using a more powerful computer.

The optimization selects a network which can shift between the production of low-protein meal (interval 60) and the production of high-protein meal and hulls (intervals 62 and 65), depending on the realization of the uncertain data.

Our industrial partner Alfa Laval has confirmed that this strategy is indeed in agreement with the consolidated industrial practice, to respond to market prices and raw material quality fluctuations. This conformity to industrial practice contributes to validate, at least qualitatively, our method and solution strategy.

It is important to underline that, since the superstructure reduction policy has been used to simplify the problem, this solution is not guaranteed to be the global optimum. The financial performances of this solution, nevertheless, appear extremely positive if compared with the performances of the optimal network under uncertainty: the expected value of the EBIT is calculated as 115.3 unit-cost/year (which corresponds to an improvement of 6.2% with respect to the network under uncertainty). It can therefore be concluded that, even though the superstructure reduction policy may drive to a local solution, this solution may still have a great value in terms of industrial decision-making, as in this case study.

Finally, this example underlines the importance of solution benchmarking, which is enabled by the incremental solution strategy integrated in our framework. Without all the information and knowledge obtained through steps 1–5, the user would not be able to benchmark the local solution obtained in step 6 against the previous one, and to understand its value for the industrial decision-making process.

Step 7. Report generation

The results are summarized in Table 10 and Fig. 7. Three different networks are identified as optimal network in deterministic condition, under uncertainty and flexible solution.

The EVPI value is quite high (19.3 unit-cost/year, 17.7% of EBIT), underlining the importance of information in the decision making. Since the majority of the uncertain data are market prices, this result underlines the importance of having good market forecast as input data for the process synthesis and design.

The other indicators confirm the outstanding financial performances of the identified flexible network, underlining how the flexibility in the product portfolio allows capturing the highest value products. This is reflected by the large value of VSS (8.8 unit-cost/year, 8.1% of the EBIT), and by the negative value of UP (−6.7 unit-cost/year). The negative UP means that, because of its flexibility, the proposed network can actually benefit from price fluctuations. From these results, the flexible network appears as the optimal decision.

5. Conclusions

A systematic framework for synthesis and design of processing networks has been proposed. By integrating methods, tools and solution strategies in a software infrastructure, the framework facilitates and optimizes the workflow required for formulation, solution and analysis of enterprise-wide optimization problems. Instead of focusing on the direct solution of the problem of decision-making under uncertainty, the framework guides the user through some preliminary steps in which the problem, the uncertainty and its consequences on the decision-making process are analyzed. The result is a comprehensive package, which contains, alongside the optimal network, a large amount of other information that constitute a valuable input for business, engineering and process management decisions. Moreover, the framework integrates an incremental solution strategy, which allows the solution of complex problems.

A network benchmark problem (NBP) has been proposed and solved to demonstrate the framework capability. It is the intention of the authors that the NBP should be treated as a benchmark problem for the development of integrated business and engineering decision-making methods and tools.

Finally, a large scale industrial problem has been studied, demonstrating the capability of the framework to manage the complexity of a real problem, subject to market, technical and raw material quality uncertainty. The solutions obtained through the application of the framework showed outstanding performance in mitigating the consequences of the uncertainty.

On the base of the presented results, our future works will focus on integrating sensitivity and ranging analysis methods and tools in our framework. Moreover, the framework will be employed for the formulation of case studies, selected from different industry segments (e.g. water and wastewater networks, production networks, biorefinery, etc.) in order to test and demonstrate its flexibility.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007–2013] under grant agreement no. 238013.

Appendix A. Proof of observation 1

Observation 1 is equivalent to state that, under the given conditions, if y^* is the optimal network under uncertainty, then a point in the uncertain domain $\theta_o \in \{\theta^{LO}; \theta^{UP}\}$ exists so that y^* is the optimal network at θ_o. In mathematical terms:

$$\text{Hypothesis}: \quad \forall y^* \quad : \quad y^* \in \{0,1\}^m; \quad \max E_\theta(z(x,y,\theta)) = E_\theta(z(x,y^*,\theta));$$

$$\text{Thesis}: \quad \exists \theta_o \quad : \quad \theta_o \in [\theta^{LO}; \theta^{UP}] \quad : \quad z(x,y^*,\theta_o) = \max[E_\theta(z(x,y,\theta_o)]$$

Proof

Consider a stochastic MILP problem (Eqs. (14)–(16)), with one uncertain parameter $\theta \in \{\theta^{LO}; \theta^{UP}\}$. For each of the n networks identified by performing the procedure described as uncertainty mapping (step 4 of the framework), a stochastic LP problem is defined.

According to the Basic Sensitivity Theorem (Fiacco & Bank, 1984), if the same base is feasible and optimal over the uncertain space, a linear relationship exists between the value of the uncertain parameter θ and the value of the objective function value z of the LP problem.

$$z(\theta) = z(\theta_o) + c \cdot (\theta - \theta_o) \tag{25}$$

where c is a constant term (often referred to as *marginal value*). For more details on Eq. (25) and the basic sensitivity theorem, including the proof, the reader is referred to (Fiacco & Bank, 1984).

The expected value of the objective function for each network is calculated as:

$$E_\theta(z_n(\theta)) = \int_{\theta_{LO}}^{\theta_{UP}} z_n(\theta) \cdot P(\theta) d\theta \tag{26}$$

where $P(\theta)$ is the probability function of θ. For uniform probability distribution and linear function $z(\theta)$ this is:

$$E_\theta(z_n(\theta)) = z_n(\theta_{MEAN}) \tag{27}$$

Table 11
Equipment list.

E-01	Stream mixer	P-01	Feed pump
E-02	Preheater/condenser	R-01	Reactor
E-03	Flash	V-01	Release valve
E-04	Final cooler		

which proves the observation, when $\theta_o = \theta_{MEAN}$. The same reasoning can be replicated for more than one uncertain data.

It is important to underline that the conditions under which this proof is obtained are quite restrictive, and correspond to particular cases in which, as shown by Eq. (27), the stochastic solution and the deterministic at mean value of the uncertain data are equivalent.

Appendix B. Use of generic process interval models

In this appendix, the formulation and solution of the generic process interval models is described through an example.

First a process will be proposed and presented through a short process description. Then the input data structure required by the generic process interval models will be explained. Finally, the calculation of the generic model data will be calculated from the process data.

Through this example, both the workflow needed for model development and the features of the development model will be highlighted.

Process description

The example process is showed in Fig. 8, and the equipment list is reported in Table 11.

The feed flow is constituted by a liquid mixture, containing a main component A and impurities B and C. The goal of the process is to reduce the content of component B in the mixture, through a decomposition reaction requiring chemical E:

$$B + 2E \leftrightarrow C + D \tag{28}$$

The reaction is performed in reactor R-01, which constitutes the core of the process.

The feed flow and the fresh chemical E are fed to mixer E-01.The reaction mixture is than brought to the reaction pressure of 4 bar by pump P-01. In E-02, the mixture is preheated to the reaction temperature of 90 °C. The heat duty required by E-02 is entirely provided by the condensation of the recycle stream, and therefore no hot utility is required.

The reaction mixture is fed to reactor R-01, where reaction (28) takes place. At design condition, 95% conversion with respect to reactant B is achieved. Because of the exothermic reaction, the temperature increases to 235 °C.

Through the pressure relief valve V-01, the reactor outlet is fed to the atmospheric flash tank E-03. The flash is adiabatic, and the flash temperature is 150 °C. The vapor outlet of flash E-03 is rich in the low boiling point components A and E. In design condition, recovery of A and E is respectively 10% and 100%. The vapor stream S10 is condensed in E-02. Recirculation of the unreacted E is not possible due to regulatory reasons, so stream S11 is sent to blow down.

The liquid outlet of flash E-03 (a mixture of A, B, C and D) is cooled in the final cooler E-04 to the final product temperature of 75 °C.

Process stream table and components molecular weights are reported in Tables 12 and 13.

Table 13
Molecular weight table.

	MW
A	180
B	48
C	42
D	42
E	18

Generic process interval data structure

The structure of the generic process interval model is reported in Fig. 9. In Table 14 the data structure is presented. For each data, the default, minimum and maximum values are shown.

Generic process interval model development

The first step of the process interval model development is the identification of process input–output. The proposed process has 3 material input and 3 material output, plus the power input to the pump. The allocation of these input–output o the category considered in the generic process interval model is reported in Table 15.

Utility and chemical use

Utility and chemical use are described by $\mu_{j,i,kk}$ and $\alpha_{i,kk}$. The former describes the amount of utility *j* which is consumed in interval *kk* as function of the mass flow of component *i* in the incoming flow. In the proposed process, 3 utilities are consumed.

Fresh E

The flow of fresh E is dosed in order to keep the desired concentration in the inlet of the reactor. As stated in the process description, an excess of 100% of E (over stoichiometry) is required to obtain the desired conversion of B. From the stoichiometry the required flow of fresh E can be calculated as:

$$R_E = 2 \cdot \frac{\gamma_E}{\gamma_B} \cdot \frac{MW_E}{MW_B} \cdot F_B^{IN} - F_E^{IN} \tag{29}$$

Since $F_E^{IN} = 0$:

$$\mu_{B,E} = 2 \cdot \frac{\gamma_E}{\gamma_B} \cdot \frac{MW_E}{MW_B} = 15 \tag{30}$$

since the utility is mixed with the process stream:

$$\alpha_{E,kk} = 1 \tag{31}$$

Table 12
Process stream table.

		s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13
Tot mass flow	(kg/m)	100.0	22.5	122.5	122.5	122.5	122.5	122.5	90.4	90.4	32.1	32.1	555.0	555.0
Mass flow A	(kg/m)	78.0	0	78.0	78.0	78.0	78.0	78.0	70.2	70.2	7.8	7.8	0	0
Mass flow B	(kg/m)	15.0	0	15.0	15.0	15.0	0.8	0.8	0.8	0.8	0	0	0	0
Mass flow C	(kg/m)	7.0	0	7.0	7.0	7.0	19.5	19.5	19.5	19.5	0	0	0	0
Mass flow D	(kg/m)	0	0	0	0	0	12.5	12.5	0	0	12.5	12.5	0	0
Mass flow E	(kg/m)	0	22.5	22.5	22.5	22.5	11.8	11.8	0	0	11.8	11.8	0	0
Mass flow CW	(kg/m)	0	0	0	0	0	0	0	0	0	0	0	555.0	555.0

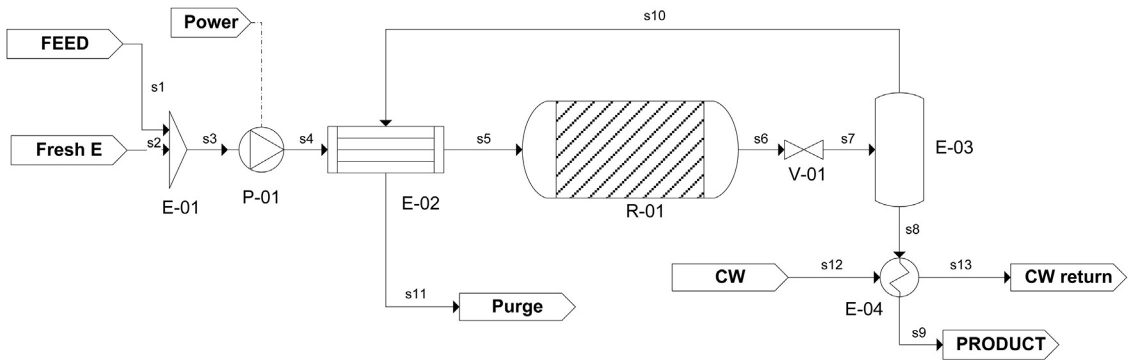

Fig. 8. Process flow diagram.

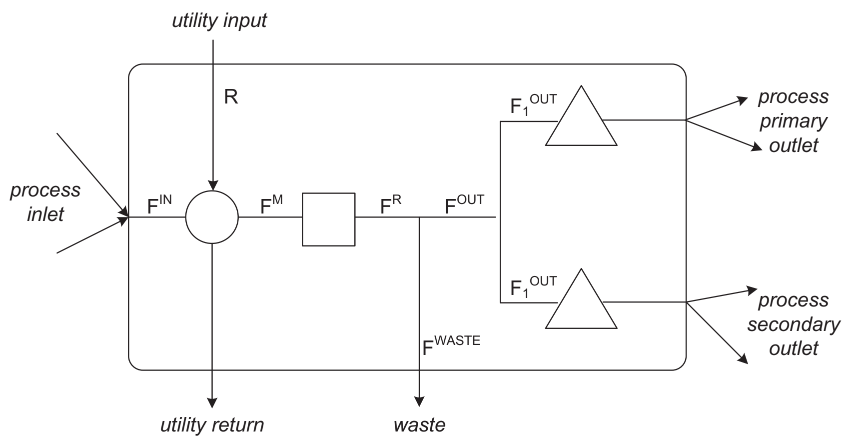

Fig. 9. Generic process interval model structure.

Table 14
Data structure of the generic interval model.

Function	Data	Description	Default value	Min value	Max value
Utility and chemicals use	$\mu_{j,i,kk}$	Amount of utility j consumed per kilogram of component i (based on flow F^{IN})	0	0	+inf
	$\alpha_{i,kk}$	Fraction of utility i mixed with the process stream (based on flow F^{IN})	0	0	1
Reaction	$\gamma_{i,kk,rr}$	Stoichiometric coefficient of component i in reaction rr (based on flow F^{M})	0	−inf	+inf
	$\theta_{react,kk,rr}$	Conversion of key reactant react in reaction rr (based on flow F^{M})	0	0	1
Waste	$SW_{i,kk}$	Fraction of component i in flow F^{R} sent to waste	0	0	1
Product separation	$Split_{i,kk}$	Fraction of component i in flow separated as primary outlet (based on flow F^{OUT})	1	0	1
Transportation	$dist_{k,kk}$	Transportation distance between interval k and interval kk (based on flow $F_1{}^{OUT}$ and $F_2{}^{OUT}$)	0	0	+inf

Table 15
Input-Output assignment.

Inlet		Outlets	
Name	Category	Name	Category
FEED	Process input	Product	Process primary outlet
Fresh E	Utility input	Purge	Waste
Power	Utility input	CW return	Utility return
CW	Utility input		

Cooling water

The amount of cooling water consumed in the final cooler is calculated solving the heat balance for the cooler:

$$R_{CW} = F_{S8} \cdot cp_{S8} \cdot \frac{T_{S8} - T_{S9}}{cp_{CW} \cdot \Delta T_{CW}} \tag{32}$$

where F_S, cp_S and T_S are respectively mass flow, heat capacity and temperature of stream s, and ΔT_{CW} is the temperature change in cooling water temperature. From this:

$$\mu_{CW,i} = \left(F_{S8} \cdot cp_{S8} \cdot \frac{T_{S8} - T_{S9}}{cp_{CW} \cdot \Delta T_{CW}} \right) \cdot \frac{1}{F^{IN}} \tag{33}$$

Since cooling water is not mixed with the process stream:

$$\alpha_{CW} = 0 \tag{34}$$

Power

The power input of pump P-01 is calculated as:

$$R_{power} = F_{S3} \cdot H_{P01} \cdot \eta_{P01} \tag{35}$$

Table 16
Utility consumption table.

$\mu_{j,i,kk}$	E	CW	Power
A	0	6.14	0.44
B	15	6.14	0.44
C	0	6.14	0.44
D	0	6.14	0.44
E	0	6.14	0.44

Table 17
Utility mix table.

α_i	
E	0
CW	1
Power	0

Table 18
Stoichiometry table.

$\gamma_{i,rr}$	Reaction 1
A	0
B	−1
C	1
D	1
E	−2

Table 19
Conversion table.

$\theta_{react,rr}$	Reaction 1
B	0.95

where W_{P01}, H_{P01} and η_{P01} are respectively power input, total dynamic head and efficiency of pump P1, and F_{S3} is total mass flow for stream S3. From this:

$$\mu_{\text{power}} = H_{P01} \cdot \eta_{P01} \tag{36}$$

Since power is not mixed with the process stream:

$$\alpha_{\text{power}} = 0 \tag{37}$$

The data structure for utility and chemical use is reported in Tables 16 and 17.

Remark

It is important to underline that the elements of the utility matrix μ can have different measurements units. As an example, data referred to fresh E are in kg/m, while the one relative to power are in kW m/kg.

Depending on the problem, it might be necessary to use different measurement units also for homogeneous data, in order to ensure good problem scaling. Particular care has to be taken in order to ensure dimensional consistence of the data employed in the model.

Reaction

The reaction function is described by use are described by $\gamma_{i,kk,rr}$ and $\theta_{\text{react},kk,rr}$. The former contains the molar stoichiometric coefficients for component *i* in reaction *rr*. The latter contains the conversion of reaction *rr* in interval *kk*, expressed with respect of the reactant *react* (key or limiting reactant).

From Eq. (28), the stoichiometry table can easily be obtained (see Table 18). As previously stated, a conversion of 95% of B is required. Component B is therefore selected as key component. The conversion table is reported in Table 19.

Table 20
Waste table.

SW_i	
A	0.1
B	0
C	0
D	0
E	1

Table 21
Separation table.

$Split_i$	
A	1
B	1
C	1
D	1
E	1

Waste

The separation of wastes from the product stream is modeled via $SW_{i,kk}$, which represents the fraction of component *i* (in flow F^R) which is separated in a waste stream. From the process description it can be seen that this corresponds to the component split factors in the flash drum. The resulting waste separation data is reported in Table 20.

Product separation

Product–product separation is intended as the use of a separation technique on the process flow, resulting in 2 process streams of different composition, named primary and secondary flow. Product-product separation is modeled via $SF_{i,kk}$, which represents the fraction of component *i* flow (in stream F^{OUT}) which goes to the primary flow.

The example process selected as example does not contain any product–product separation, therefore the elements of $SF_{i,kk}$ are at their default value (equal to 1, Table 21) meaning that no separation is occurring and all mass flow goes to primary outlet.

Transportation

Transportation cost is modeled via $dist_{k,kk}$, which represents the geographical distance between interval *k* and interval *kk*. Transportation cost is not considered in the selected example, therefore all the elements of $dist_{k,kk}$ are equal to the default value of zero.

Appendix C. Network benchmark problem

In this appendix, the network benchmark problem is described by presenting problem equation and data.

The network benchmark problem is a network synthesis problem based on a superstructure composed of 12 intervals (Fig. 6), 4 process components (C-1, C-2, C-3, C-4) and one utility component (U-1). The objective function is the maximization the Earning Before Interests and Tax (EBIT).

The NBP can be formulated as a MILP problem. Hereinafter, both the mathematical formulation (Eqs. (38)–(56)) and the complete set of data (Tables 22–32) are reported, to allow the reader to reproduce the problem and its solution.

$$\max EBIT = \sum_{i,kk}(P3_{i,kk} \cdot F^{out}_{i,kk}) - \sum_{i,kk}(P2_{i,kk} \cdot R_{i,kk}) - \sum_{i,kk}(P1_{i,kk} \cdot F^{out}_{i,kk}) - W^{Price} \cdot \sum_{i,kk}(F^{R}_{i,kk} \cdot SW_{i,kk}) - CAPEX/t \tag{38}$$

Please cite this article in press as: Quaglia, A., et al. A systematic framework for enterprise-wide optimization: Synthesis and design of processing networks under uncertainty. *Computers and Chemical Engineering* (2013), http://dx.doi.org/10.1016/j.compchemeng.2013.03.018

Table 22
NBP data – superstructure matrix S.

S	I-1	I-2	II-1	II-2	II-3	III-1	III-2	IV-1	V-1	V-2	V-3	V-4
I-1	0	0	1	1	1	0	0	0	0	0	0	0
I-2	0	0	1	1	1	0	0	0	0	0	0	0
II-1	0	0	0	0	0	1	1	0	0	0	0	0
II-2	0	0	0	0	0	1	1	0	0	0	0	0
II-3	0	0	0	0	0	0	1	1	0	0	0	1
III-1	0	0	0	0	0	0	0	0	1	1	0	0
III-2	0	0	0	0	0	0	0	0	0	0	1	0
IV-1	0	0	0	0	0	0	0	0	0	0	1	0

Table 23
NBP data – primary outlet matrix SP.

SP	I-1	I-2	II-1	II-2	II-3	III-1	III-2	IV-1	V-1	V-2	V-3	V-4
I-1	0	0	1	1	1	0	0	0	0	0	0	0
I-2	0	0	1	1	1	0	0	0	0	0	0	0
II-1	0	0	0	0	0	1	1	0	0	0	0	0
II-2	0	0	0	0	0	1	1	0	0	0	0	0
II-3	0	0	0	0	0	0	1	0	0	0	0	0
III-1	0	0	0	0	0	0	0	0	1	0	0	0
III-2	0	0	0	0	0	0	0	0	0	0	1	0
IV-1	0	0	0	0	0	0	0	0	0	0	1	0

Table 24
NBP data – split factor SF.

$SF_{i,kk}$	I-1	I-2	II-1	II-2	II-3	III-1	III-2	IV-1	V-1	V-2	V-3	V-4
C-1	1.00	1.00	1.00	1.00	0.95	0.70	1.00	1.00	1.00	1.00	1.00	1.00
C-2	1.00	1.00	1.00	1.00	0.99	0.10	1.00	1.00	1.00	1.00	1.00	1.00
C-3	1.00	1.00	1.00	1.00	0.50	0.00	1.00	1.00	1.00	1.00	1.00	1.00
C-4	1.00	1.00	1.00	1.00	0.50	0.00	1.00	1.00	1.00	1.00	1.00	1.00
U-1	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 25
NBP data – utility matrix μ.

$\mu_{i,k,kk}$		II-1	II-2	II-3	III-1	III-2	IV-1
U-1	A	0.4	0.0	0.5	0.9	0.0	0.0
U-1	B	0.4	1.5	0.5	0.0	2.5	0.0
U-1	C	0.4	0.0	0.5	0.0	0.0	0.0
U-1	D	0.4	0.0	0.5	0.0	0.0	0.0

Table 26
NBP data – stoichiometry table γ.

$\gamma_{i,kk,rr}$		r1	r2
C-2	II-2	−1	
U-1	II-2	−2	
C-3	II-2	1	
C-4	II-2	1	
C-3	III-2		−1
C-4	III-2		1

Table 27
NBP data – conversion table θ.

$\theta_{react,kk,rr}$		r1	r2
C-2	II-2	0.95	
C-3	III-2		0.90

Table 28
NBP data–Utility mix matrix α.

$\alpha_{i,kk}$	U-1
II-1	0
II-2	1
II-3	0
III-1	0
III-2	0
IV-1	0

Table 29
NBP data – waste matrix SW.

$SW_{i,kk}$	II-1	II-2	II-3	III-1	III-2	IV-1
C-1	0.0	0.05	0.0	0.0	0.0	0.0
C-2	0.9	0.0	0.2	0.0	0.4	0.8
C-3	0.0	0.0	0.0	0.2	0.0	0.8
C-4	0.0	0.0	0.0	0.0	0.0	0.0
U-1	0.0	1.0	0.0	0.0	0.0	0.0

Table 30
NBP data – capital cost data.

Capex	II-1	II-2	II-3	III-1	III-2	IV-1
$\alpha^L_{j,kk}$	12,500	7500	8000	17,500	8000	5000
$\beta^L_{j,kk}$	0.55	0.65	0.65	0.66	0.5	0.6
F^{MAX}_{kk}	200	125	200	190	140	150

Table 31
NBP data – raw material component flow.

$\Phi_{i,kk}$	1	2
C-1	78	75
C-2	15	20
C-3	7	5
C-4	0	0
U-1	0	0

Table 32
NBP data – miscellaneous data.

Data	Value
t	10 years
M	1E5
Wprice	15

$$F^M_{i,kk} = \sum_k (F_{i,k,kk}) + \alpha_{i,kk} \cdot R_{i,kk} \tag{39}$$

$$R_{i,kk} = \mu_{i,kk} \cdot \sum_{i,k} (F_{i,k,kk}) \tag{40}$$

$$F^R_{i,kk} = F^M_{i,kk} + \sum_{rr,react} (\gamma_{i,kk,rr} \cdot \theta_{react,kk,rr} \cdot F^M_{react,kk}) \tag{41}$$

$$F^{out}_{i,kk} = F^R_{i,kk} \cdot (1 - SW_{i,kk}) \quad \text{for} \quad kk \quad \text{is processes} \tag{42}$$

$$F^{out}_{i,kk} = \phi_{i,kk} \quad \text{for} \quad kk \quad \text{is raw materials} \tag{43}$$

$$F^{WASTE}_{i,kk} = F^R_{i,kk} - F^{out}_{i,kk} \tag{44}$$

$$F^{out1}_{i,kk} = F^{out}_{i,kk} \cdot SF_{i,kk} \tag{45}$$

$$F^{out2}_{i,kk} = F^{out}_{i,kk} \cdot (1 - SF_{i,kk}) \tag{46}$$

$$F^1_{i,k,kk} \le F^{out1}_{i,kk} \cdot S_p \tag{47}$$

$$F^2_{i,k,kk} \le F^{out2}_{i,kk} \cdot (S - S_p) \tag{48}$$

$$\sum_k F^1_{i,k,kk} = F^{out1}_{i,kk} \tag{49}$$

$$\sum_k F^2_{i,k,kk} = F^{out2}_{i,kk} \tag{50}$$

Please cite this article in press as: Quaglia, A., et al. A systematic framework for enterprise-wide optimization: Synthesis and design of processing networks under uncertainty. *Computers and Chemical Engineering* (2013), http://dx.doi.org/10.1016/j.compchemeng.2013.03.018

$$\sum_i F^R_{i,kk} \le F^{\mathrm{MAX}}_{kk} \tag{51}$$

$$CAPEX = \sum_{kk} \left[\sum_j (\alpha^L_{j,kk} \cdot w_{j,kk} + \beta^L_{j,kk} \cdot Q_{j,kk}) \right] \tag{52}$$

$$\sum_i F^R_{i,kk} = \sum_j Q_{j,kk} \tag{53}$$

$$Q^o_{j,kk} \cdot w_{j,kk} \le Q_{j,kk} \le Q^o_{j+1,kk} \cdot w_{j,kk} \tag{54}$$

$$\sum_j w_{j,kk} = 1 \tag{55}$$

$$F^R_{i,kk} \le M \cdot y_{kk} \tag{56}$$

Additional logical constraints are added to exclude redundant solutions from the search space (e.g. no stream split condition).

$$y_{I-1} + y_{I-2} \le 1 \tag{57}$$

$$y_{II-1} + y_{II-2} + y_{II-3} \le 1 \tag{58}$$

$$y_{III-1} + y_{III-2} \le 1 \tag{59}$$

The NBP formulation is based on the generic process interval model structure proposed in (Quaglia et al., 2012a, 2012b), with some minor modifications. An explicit formulation is used to model product–product separation (Eqs. (45)–(50)). Piecewise linearization strategy is used for the calculation of the capital cost with respect to the process throughput (Eqs. (52)–(55)).

If 10 intervals are used for the piecewise linearization, the NBP problem is constituted by 2260 constraints and 120 binary variables.

References

Birge, J. K., & Louveaux, F. (1999). *Introduction to stochastic programming (Springer series in operations research and financial engineering)*. Springer: New York.

Dua, V., & Pistikopoulos, E. N. (1998). Optimization techniques for process synthesis and material design under uncertainty. *Chemical Engineering Research and Design*, *76*, 408–416.

Erdirik-Dogan, M., Grossmann, E., & Wassick, I. J. (2007). A bi-level decomposition scheme for the integration of planning and scheduling in parallel multi-product batch reactors. *Computer Aided Chemical Engineering*, *24*, 625–630.

Erdirik-Dogan, M., & Grossmann, I. E. (2008). Simultaneous planning and scheduling of single-stage multi-product continuous plants with parallel lines. *Computers and Chemical Engineering*, *32*, 2664–2683.

Fiacco, A., & Bank, B. (1984). *Sensitivity stability and parametric analysis*. Amsterdam: North-Holland.

Floudas, C. A. (1995). *Nonlinear and mixed-integer optimization: Fundamentals and applications*. New York: Oxford University Press.

Grossmann, I. (2005). Enterprise-wide optimization: A new frontier in process systems engineering. *AICHE Journal*, *51*, 1846–1857.

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, *81*, 23–69.

IBM Corp. (2009). IBM ILOG CPLEX V 12.1. IBM Corp.

Ierapetritou, M. G., Acevedo, J., & Pistikopoulos, E. N. (1996). An optimization approach for process engineering problems under uncertainty. *Computers and Chemical Engineering*, *20*, 703–709.

Ierapetritou, M. G., Pistikopoulos, E. N., & Floudas, C. A. (1996). Operational planning under uncertainty. *Computers and Chemical Engineering*, *20*, 1499–1516.

Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics – Simulation and Computation*, *11*, 311–334.

Karuppiah, R., & Grossmann, I. E. (2006). Global optimization for the synthesis of integrated water systems in chemical processes. *Computers and Chemical Engineering*, *30*, 650–673.

Karuppiah, R., & Grossmann, I. E. (2008). Global optimization of multiscenario mixed integer nonlinear programming models arising in the synthesis of integrated water networks under uncertainty. *Computers and Chemical Engineering*, *32*, 145–160.

Kopanos, G. M., Puigjaner, L., & Georgiadis, M. C. (2009). A bi-level decomposition methodology for scheduling batch chemical production facilities. *Computer Aided Chemical Engineering*, *27*, 681–686.

Paules, G. E., IV, & Floudas, C. A. (1992). Stochastic programming in process synthesis: A two-stage model with MINLP recourse for multiperiod heat-integrated distillation sequences. *Computers and Chemical Engineering*, *16*, 189–210.

Project Management Institute. (2008). *A guide to the project management body of knowledge* (4th ed.). Newton Square, PA: Project Management Institute.

Quaglia, A., Sarup, B., Sin, G., & Gani, R. (2012a). Integrated business and engineering framework for synthesis and design of enterprise-wide processing networks. *Computers and Chemical Engineering*, *38*, 213–223.

Quaglia, A., Sarup, B., Sin, G., & Gani, R. (2012b). Synthesis and design of processing networks: Stochastic formulation and solution. *Computer Aided Chemical Engineering*, *30*, 467–471.

Sahinidis, N. V. (2004). Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering*, *28*, 971–983.

Shapiro, A. (2012). Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research*, *219*, 719–726.

Sin, G., Lantz, A. E., & Gernaey, K. V. (2009). Sensitivity analysis of non-linear dynamic models: Prioritizing experimental research. *Computer Aided Chemical Engineering*, *27*, 267–272.

Terrazas-Moreno, S., & Grossmann, I. E. (2011). A multiscale decomposition method for the optimal planning and scheduling of multi-site continuous multiproduct plants. *Chemical Engineering Science*, *66*, 4307–4318.

Varma, V. A., Reklaitis, G. V., Blau, G. E., & Pekny, J. F. (2007). Enterprise-wide modeling & optimization—An overview of emerging research challenges and opportunities. *Computers and Chemical Engineering*, *31*, 692–711.